

Multicolor Sunflowers

Dhruv Mubayi* Lujia Wang†

December 15, 2015

Abstract

A sunflower is a collection of distinct sets such that the intersection of any two of them is the same as the common intersection C of all of them, and $|C|$ is smaller than each of the sets. A longstanding conjecture due to Erdős and Szemerédi states that the maximum size of a family of subsets of $[n]$ that contains no sunflower of fixed size $k > 2$ is exponentially smaller than 2^n as $n \rightarrow \infty$. We consider this problem for multiple families. In particular, we obtain sharp or almost sharp bounds on the sum and product of k families of subsets of $[n]$ that together contain no sunflower of size k with one set from each family. For the sum, we prove that the maximum is

$$(k-1)2^n + 1 + \sum_{s=n-k+2}^n \binom{n}{s}$$

for all $n \geq k \geq 3$, and for the $k = 3$ case of the product, we prove that it is between

$$\left(\frac{1}{8} + o(1)\right) 2^{3n} \quad \text{and} \quad (0.13075 + o(1)) 2^{3n}.$$

1 Introduction

Throughout the paper, we write $[n] = \{1, \dots, n\}$, $2^{[n]} = \{S : S \subset [n]\}$ and $\binom{[n]}{s} = \{S : S \subset [n], |S| = s\}$. A *sunflower* (or *strong Δ -system*) with k petals is a collection of k sets $\mathcal{S} = \{S_1, \dots, S_k\}$ such that $S_i \cap S_j = C$ for all $i \neq j$, and $S_i \setminus C \neq \emptyset$ for all $i \in [k]$. The common intersection C is called the *core* of the sunflower and the sets $S_i \setminus C$ are called the *petals*. In 1960, Erdős and Rado [9] proved a fundamental result regarding the existence of sunflowers in a large family of sets of uniform size, which is now referred to as the *sunflower lemma*. It states that if \mathcal{A} is a family of sets of size s with $|\mathcal{A}| > s!(k-1)^s$, then \mathcal{A} contains a sunflower with k petals. Later in 1978, Erdős and Szemerédi [10] gave the following upper bound when the underlying set has n elements.

*Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL 60607, Email:mubayi@uic.edu. Research partially supported by NSF grants DMS-0969092 and DMS-1300138

†Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL 60607, Email:lwang203@uic.edu.

Theorem 1 (Erdős, Szemerédi [10]). *There exists a constant c such that if $\mathcal{A} \subset 2^{[n]}$ with $|\mathcal{A}| > 2^{n-c\sqrt{n}}$ then \mathcal{A} contains a sunflower with 3 petals.*

In the same paper, they conjectured that for n sufficiently large, the maximum number of sets in a family $\mathcal{A} \subset 2^{[n]}$ with no sunflowers with three petals is at most $(2 - \epsilon)^n$ for some absolute constant $\epsilon > 0$. This conjecture remains open, and is closely related to the algorithmic problem of matrix multiplication, see [1].

Similar problems have been studied for systems of sets where only the size (rather than the actual set) of pairwise intersections is fixed. A *weak Δ -system* of size k is a collection of k sets $\mathcal{S} = \{S_1, \dots, S_k\}$ such that $|S_i \cap S_j| = |S_1 \cap S_2|$ for all $i \neq j$. Thus, a sunflower is a weak Δ -system but not vice versa. In 1973, Deza [7] gave the criterion for a weak Δ -system to be a sunflower: If \mathcal{F} is an s -uniform weak Δ -system with $|\mathcal{F}| > s^2 - s + 1$, then \mathcal{F} is a sunflower. The lower bound can be achieved only if the projective plane $PG(2, s)$ exists. This was shown by van Lint [18] later in the same year. Erdős posed the problem of determining the largest size of a family $\mathcal{A} \subset 2^{[n]}$ that contains no weak Δ -system of a fixed size. The problem was solved by Frankl and Rödl [11] in 1987. They proved that given $k \geq 3$, there exists a constant $\epsilon = \epsilon(k)$ so that for every $\mathcal{A} \subset 2^{[n]}$ with $|\mathcal{A}| > (2 - \epsilon)^n$, \mathcal{A} contains a weak Δ -system of size k .

A natural way to generalize problems in extremal set theory is to consider versions for multiple families or so-called multicolor or cross-intersecting problems. Beginning with the famous Erdős-Ko-Rado theorem [8], which states that an intersecting family of k -element subsets of $[n]$ has size at most $\binom{n-1}{k-1}$, provided $n \geq 2k$, several generalizations were proved for multiple families that are cross-intersecting. In particular, Hilton [14] showed in 1977 that if t families $\mathcal{A}_1, \dots, \mathcal{A}_t \subset \binom{[n]}{k}$ are cross intersecting (meaning that $A_i \cap A_j \neq \emptyset$ for all $(A_i, A_j) \in \mathcal{A}_i \times \mathcal{A}_j$) and if $n/k \leq t$, then $\sum_{i=1}^t |\mathcal{A}_i| \leq t \binom{n-1}{k-1}$. On the other hand, results of Pyber [21] in 1986, that were later slightly refined by Matsumoto and Tokushige [19] and Bey [2], showed that if two families $\mathcal{A} \subset \binom{[n]}{k}$, $\mathcal{B} \subset \binom{[n]}{l}$ are cross-intersecting and $n \geq \max\{2k, 2l\}$, then $|\mathcal{A}||\mathcal{B}| \leq \binom{n-1}{k-1} \binom{n-1}{l-1}$. These are the first results about bounds on sums and products of the size of cross-intersecting families. More general problems were considered recently, for example for cross t -intersecting families (i.e. pair of sets from distinct families have intersection of size at least t) and r -cross intersecting families (any r sets have a nonempty intersection where each set is picked from a distinct family) and labeled crossing intersecting families, see [4, 12, 13]. A more systematic study of multicolored extremal problems (with respect to the sum of the sizes of the families) was begun by Keevash, Saks, Sudakov, and Verstraëte [16], and continued in [3, 17]. Cross-intersecting versions of Erdős' problem on weak Δ -systems mentioned above (for the product of the size of two families) were proved by Frankl and Rödl [11] and by the first author and Rödl [20].

In this note, we consider multicolor versions of sunflower theorems. Quite surprisingly, these basic questions appear not to have been studied in the literature.

Definition 2. Given families of sets $\mathcal{A}_i \subset 2^{[n]}$ for $i = 1, \dots, k$, a *multicolor sunflower* with k petals is a collection of sets $A_i \in \mathcal{A}_i$, $i = 1, \dots, k$, such that $A_i \cap A_j = C$ for all $i \neq j$, and $A_i \setminus C \neq \emptyset$, for all $i \in [k]$. Say that $\mathcal{A}_1, \dots, \mathcal{A}_k$ is sunflower-free if it contains no multicolor

sunflower with k petals.

For any k families that are sunflower-free, the problem of upper bounding the size of any single family is uninteresting, since there is no restriction on a particular family. So we are interested in the sum and product of the sizes of these families.

Given integers n and k , let

$$\mathcal{F}(n, k) = \{ \{ \mathcal{A}_i \}_{i=1}^k : \mathcal{A}_i \subset 2^{[n]} \text{ for } i \in [k] \text{ and } \mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_k \text{ is sunflower-free} \}.$$

We define

$$S(n, k) := \max_{\{ \mathcal{A}_i \}_{i=1}^k \in \mathcal{F}(n, k)} \sum_{i=1}^k |\mathcal{A}_i|,$$

and

$$P(n, k) := \max_{\{ \mathcal{A}_i \}_{i=1}^k \in \mathcal{F}(n, k)} \prod_{i=1}^k |\mathcal{A}_i|.$$

2 Main Results

Our two main results are sharp or nearly sharp estimates on $S(n, k)$ and $P(n, 3)$. By Theorem 1 we obtain that

$$S(n, 3) \leq 2 \cdot 2^n + 2^{n-c\sqrt{n}}.$$

Indeed, if $|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}|$ is larger than the RHS above then $|\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}| > 2^{n-c\sqrt{n}}$ by the pigeonhole principle and we find a sunflower in the intersection which contains a multicolor sunflower. Our first result reduces the term $2^{n-c\sqrt{n}}$ to obtain an exact result.

Theorem 3. *For $n \geq k \geq 3$*

$$S(n, k) = (k-1)2^n + 1 + \sum_{s=n-k+2}^n \binom{n}{s}.$$

The problem of determining $P(n, k)$ seems more difficult than that of determining $S(n, k)$. Our bounds for general k are quite far apart, but in the case $k = 3$ we can refine our argument to obtain a better bound.

Theorem 4.

$$\left(\frac{1}{8} + o(1) \right) 2^{3n} \leq P(n, 3) \leq (0.13075 + o(1)) 2^{3n}.$$

We conjecture that the lower bound is tight.

Conjecture 1. For each fixed $k \geq 3$,

$$P(n, k) = \left(\frac{1}{8} + o(1) \right) 2^{kn}.$$

In the next two subsections we give the proofs of Theorems 3 and 4.

2.1 Sums

In order to prove Theorem 3, we first deal with s -uniform families and prove a stronger result. Given a multicolor sunflower \mathcal{S} , define its *core size* to be $c(\mathcal{S}) = |C|$.

Lemma 5. Given integers $s \geq 1$ and c with $0 \leq c \leq s - 1$, let n be an integer such that $n \geq c + k(s - c)$. For $i = 1, \dots, k$, let $\mathcal{A}_i \subset \binom{[n]}{s}$ such that $\{\mathcal{A}_i\}_{i=1}^k$ contains no multicolor sunflower with k petals and core size c . Then

$$\sum_{i=1}^k |\mathcal{A}_i| \leq (k - 1) \binom{n}{s}.$$

Furthermore, this bound is tight.

Proof. Randomly take an ordered partition of $[n]$ into $k + 2$ parts X_1, X_2, \dots, X_{k+2} such that $|X_1| = n - (c + k(s - c))$, $|X_2| = c$, and $|X_i| = s - c$ for $i = 3, \dots, k + 2$, with uniform probability for each partition. For each partition, construct the bipartite graph

$$G = (\{\mathcal{A}_i : i = 1, \dots, k\} \cup \{X_2 \cup X_j : j \in [3, k + 2]\}, E)$$

where a pair $\{\mathcal{A}_i, X_2 \cup X_j\} \in E$ if and only if $X_2 \cup X_j \in \mathcal{A}_i$. If there exists a perfect matching in G , then we will get a multicolor sunflower with k petals and core size c , since X_2 will be the core. This shows that G has matching number at most $k - 1$. Then König's theorem implies that the random variable $|E(G)|$ satisfies

$$|E(G)| \leq (k - 1)k. \tag{1}$$

Another way to count the edges of G is through the following expression:

$$|E(G)| = \sum_{i=1}^k \sum_{j=3}^{k+2} \chi_{\{X_2 \cup X_j \in \mathcal{A}_i\}}, \tag{2}$$

where χ_A is the characteristic function of the event A . Taking expectations and using (1) we obtain

$$\mathbb{E} \left(\sum_{i=1}^k \sum_{j=3}^{k+2} \chi_{\{X_2 \cup X_j \in \mathcal{A}_i\}} \right) \leq (k - 1)k.$$

By linearity of expectation,

$$\mathbb{E} \left(\sum_{i=1}^k \sum_{j=3}^{k+2} \chi_{\{X_2 \cup X_j \in \mathcal{A}_i\}} \right) = \sum_{i=1}^k \sum_{j=3}^{k+2} \mathbb{P}(X_2 \cup X_j \in \mathcal{A}_i) = \sum_{i=1}^k \sum_{j=3}^{k+2} \sum_{A \in \mathcal{A}_i} \mathbb{P}(A = X_2 \cup X_j).$$

The probability that a set A is partitioned as $X_2 \cup X_j$ is the same as the probability that A is partitioned into two ordered parts of sizes c and $s - c$, and $[n] \setminus A$ has an ordered partitioned into k parts with one of the parts of size $n - (c + k(s - c))$ and $k - 1$ of them of size $s - c$. Hence for any $A \in \mathcal{A}_i$,

$$\begin{aligned} \mathbb{P}(A = X_2 \cup X_j) &= \frac{\binom{|A|}{c} \binom{n-|A|}{n-(c+k(s-c))} \prod_{i=1}^{k-1} \binom{(k-i)(s-c)}{s-c}}{\binom{n}{c+k(s-c)} \binom{c+k(s-c)}{c} \prod_{i=0}^{k-1} \binom{(k-i)(s-c)}{s-c}} \\ &= \frac{\binom{s}{c} \binom{n-s}{n-(c+k(s-c))} \prod_{i=1}^{k-1} \binom{(k-i)(s-c)}{s-c}}{\binom{n}{c+k(s-c)} \binom{c+k(s-c)}{c} \binom{k(s-c)}{s-c} \prod_{i=1}^{k-1} \binom{(k-i)(s-c)}{s-c}} \\ &= \frac{1}{\binom{n}{s}}. \end{aligned}$$

So we have

$$\mathbb{E} \left(\sum_{i=1}^k \sum_{j=3}^{k+2} \chi_{\{X_2 \cup X_j \in \mathcal{A}_i\}} \right) = \sum_{i=1}^k \sum_{j=3}^{k+2} \sum_{A \in \mathcal{A}_i} \frac{1}{\binom{n}{s}} = \sum_{i=1}^k |\mathcal{A}_i| \frac{k}{\binom{n}{s}}.$$

Hence by (2),

$$\sum_{i=1}^k |\mathcal{A}_i| \leq (k-1) \binom{n}{s}.$$

The bound shown above is tight, since we can take $\mathcal{A}_1 = \mathcal{A}_2 = \dots = \mathcal{A}_{k-1} = \binom{[n]}{s}$, and $\mathcal{A}_k = \emptyset$. \square

Now we use this lemma to prove Theorem 3.

Proof of Theorem 3. Recall that $n \geq k \geq 3$ and we are to show that

$$S(n, k) = (k-1)2^n + 1 + \sum_{s=n-k+2}^n \binom{n}{s}.$$

To see the upper bound, given families $\{\mathcal{A}_i\}_{i=1}^k \in \mathcal{F}(n, k)$, we define $\mathcal{A}_{i,s} = \mathcal{A}_i \cap \binom{[n]}{s}$ for each $i \in [k]$ and integer $s \in [0, n]$. This gives a partition of each family \mathcal{A}_i into $n+1$ subfamilies. Since families $\{\mathcal{A}_i\}_{i=1}^k$ contain no multicolor sunflowers with k petals, neither do $\{\mathcal{A}_{i,s}\}_{i=1}^k$ for all $s \in [0, n]$. Now, for each $s = 1, 2, \dots, n-k+1$ let

$$c = \max \left\{ 0, \frac{n - ks}{1 - k} \right\}.$$

Then $0 \leq c \leq s - 1$, and $n \geq c + k(s - c)$. Therefore, by Lemma 5, for $1 \leq s \leq n - k + 1$,

$$\sum_{i=1}^k |\mathcal{A}_{i,s}| \leq (k-1) \binom{n}{s}.$$

For $s > n - k + 1$, notice that a trivial bound for this sum is $k \binom{n}{s}$. So we get,

$$\begin{aligned} \sum_{i=1}^k |\mathcal{A}_i| &= \sum_{i=1}^k \sum_{s=0}^n |\mathcal{A}_{i,s}| \\ &= \sum_{s=0}^n \sum_{i=1}^k |\mathcal{A}_{i,s}| \\ &= \sum_{i=1}^k |\mathcal{A}_{i,0}| + \sum_{s=1}^{n-k+1} \sum_{i=1}^k |\mathcal{A}_{i,s}| + \sum_{s=n-k+2}^n \sum_{i=1}^k |\mathcal{A}_{i,s}| \\ &\leq k \binom{n}{0} + \sum_{s=1}^{n-k+1} (k-1) \binom{n}{s} + \sum_{s=n-k+2}^n k \binom{n}{s} \\ &\leq \sum_{s=0}^n (k-1) \binom{n}{s} + \binom{n}{0} + \sum_{s=n-k+2}^n \binom{n}{s} \\ &= (k-1)2^n + 1 + \sum_{s=n-k+2}^n \binom{n}{s}. \end{aligned}$$

The lower bound is obtained by the following example: $\mathcal{A}_i = 2^{[n]}$ for $i = 1 \dots, k-1$ and $\mathcal{A}_k = \{\emptyset\} \cup \{S \subset [n] : |S| \geq n - k + 2\}$. To see that $\{\mathcal{A}_i\}_{i=1}^k$ contains no multicolor sunflower, notice that any multicolor sunflower uses a set from \mathcal{A}_k . The empty set does not lie in any sunflowers. So if a set of size at least $n - k + 2$ appeared in a sunflower with k petals, it requires at least $k - 1$ other points to form such a sunflower, but then the total number of points in this sunflower is at least $n + 1$, a contradiction. \square

2.2 Products

From the bound on the sum of the families that do not contain a multicolor sunflower, we deduce the following bound on the product by using AM-GM inequality.

Corollary 6. Fix $k \geq 3$. As $n \rightarrow \infty$,

$$\left(\frac{1}{8} + o(1)\right) 2^{kn} \leq P(n, k) \leq \left(\left(\frac{k-1}{k}\right)^k + o(1)\right) 2^{kn}.$$

Proof. The upper bound follows from Theorem 3 and the AM-GM inequality,

$$\prod_{i=1}^k |\mathcal{A}_i| \leq \left(\frac{\sum_{i=1}^k |\mathcal{A}_i|}{k}\right)^k \leq \left((1 + o(1)) \frac{(k-1)2^n}{k}\right)^k = (1 + o(1)) \left(\frac{k-1}{k}\right)^k 2^{kn}.$$

For the lower bound, we take

$$\mathcal{A}_1 = \mathcal{A}_2 = \{S \subset [n] : 1 \in S \text{ or } |S| \geq n - 1\},$$

$$\mathcal{A}_3 = \{S \subset [n] : 1 \notin S \text{ or } |S| \geq n - 1\},$$

and $\mathcal{A}_4 = \mathcal{A}_5 = \dots = \mathcal{A}_k = 2^{[n]}$. A multicolor sunflower with k petals must use three sets from $\mathcal{A}_1, \mathcal{A}_2$, and \mathcal{A}_3 , call them A_1, A_2, A_3 respectively. These three sets form a multicolor sunflower with three petals. If any of these sets is of size at least $n - 1$, then it will be impossible to form a 3-petal sunflower with the other two sets. So by their definitions, we have $1 \in A_1 \cap A_2$, but $1 \notin A_3$, which implies $A_1 \cap A_2 \neq A_1 \cap A_3$, a contradiction. So the families $\{\mathcal{A}_i\}_{i=1}^k$ contain no multicolor sunflowers with k petals. The sizes of these families are $|\mathcal{A}_1| = |\mathcal{A}_2| = |\mathcal{A}_3| = (\frac{1}{2} + o(1))2^n$, and $|\mathcal{A}_i| = 2^n$ for $i \geq 4$. Thus,

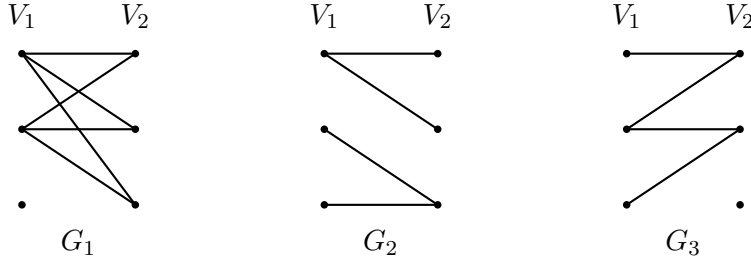
$$\prod_{i=1}^k |\mathcal{A}_i| = \left(\frac{1}{8} + o(1)\right) 2^{kn}.$$

□

For any positive integer k we have $(\frac{k-1}{k})^k < 1/e$, so Corollary 6 implies the upper bound $(1/e + o(1))2^{kn}$ for all $k \geq 3$. For $k = 3$, we will improve the factor in the upper bound from $(2/3)^3 = 0.29629\dots$ to approximately 0.131, which is quite close to our conjectured value of 0.125. We need the following lemma.

Lemma 7. *Let $G = (V_1 \cup V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = 3$ and $d(v) \leq 2$ for all $v \in V_2$. If the maximum size of a matching in G is at most two, then G is a subgraph of one of the following three graphs.*

- G_1 : a copy of $K_{2,3}$ with the part of size two in V_1 and the part of size three in V_2
- G_2 : two vertex disjoint copies of the path with two edges
- G_3 : a path with four edges whose endpoints are in V_1



Proof. By König's theorem, the minimum vertex cover of G has size at most two. Suppose G has a vertex that covers all the edges. Then G is a subgraph of either a $K_{1,3}$ whose

degree three vertex is in V_1 or a $K_{1,2}$ whose degree two vertex is in V_2 . Both of these are subgraphs of G_1 .

Now we may assume that a minimum vertex cover S of G has size two. If $S \subset V_1$, then G is a subgraph of G_1 . Next we assume that $S = \{u, v\}$ with $u \in V_1, v \in V_2$. If $d(u) = 3$, then $d(v) \leq 2$ implies that $G \subset G_1$. If $d(u) = 2$ and $uv \in E(G)$, then $G \subset G_1$. If $d(u) = 2$ and $uv \notin E(G)$, then $G \subset G_2$. If $d(u) = 1$, then clearly $G \subset G_3$. The remaining case is that $S \subset V_2$, and it is obvious that $G \subset G_1$ or $G \subset G_3$. \square

We now have the necessary ingredients to prove Theorem 4.

Proof of Theorem 4. Recall that $n \geq k \geq 3$ and we are to show that

$$\left(\frac{1}{8} + o(1)\right) 2^{3n} \leq P(n, 3) \leq (0.13075 + o(1)) 2^{3n}.$$

The lower bound follows from Corollary 6. The upper bound is proved using a similar idea as the proof of Lemma 5, although the graph statistic considered is more complicated. First notice that given families $\mathcal{A}_i \subset 2^{[n]}, i = 1, 2, 3$ that are in $\mathcal{F}(n, 3)$, such that $\prod_{i=1}^3 |\mathcal{A}_i|$ is maximized, we may assume that the common part of the three families $\bigcap_{i=1}^3 \mathcal{A}_i = \emptyset$. To see this, let $\mathcal{A}_c = \bigcap_{i=1}^3 \mathcal{A}_i$. By Theorem 1, $|\mathcal{A}_c| \leq 2^{n-c\sqrt{n}} = o(2^n)$, otherwise a 3-petal multicolor sunflower exists. Notice also that from the lower bound, $|\mathcal{A}_i| = \Theta(2^n)$ for all $i \in [3]$, we have

$$\prod_{i=1}^3 |\mathcal{A}_i| = \prod_{i=1}^3 (|\mathcal{A}_i \setminus \mathcal{A}_c| + |\mathcal{A}_c|) = \prod_{i=1}^3 (|\mathcal{A}_i \setminus \mathcal{A}_c| + o(2^n)) = \prod_{i=1}^3 |\mathcal{A}_i \setminus \mathcal{A}_c| + o(2^{3n}).$$

Hence, it suffices to show a bound on $\prod_{i=1}^3 |\mathcal{A}_i \setminus \mathcal{A}_c|$.

We uniformly take an ordered partition of $[n] = X_1 \cup X_2 \cup X_3 \cup X_4$ at random such that the parts X_2, X_3 and X_4 are nonempty. So there are

$$p(n) = 4^n - 3 \cdot 3^n + 3 \cdot 2^n - 1 = 4^n + O(3^n)$$

such partitions, each is chosen with probability $1/p(n)$. Again, we construct a bipartite graph $G = (V_1 \cup V_2, E)$, where $V_1 = \{\mathcal{A}_i : i = 1, 2, 3\}$ and $V_2 = \{X_1 \cup X_j : j = 2, 3, 4\}$, and a pair $\{\mathcal{A}_i, X_1 \cup X_j\} \in E$ if and only if $X_1 \cup X_j \in \mathcal{A}_i$. A perfect matching in G gives rise to a multicolor sunflower with three petals. Hence the maximum size of a matching in G is at most two. Moreover, since $\mathcal{A}_c = \emptyset$, the degrees of vertices in V_2 are at most two. We may now apply Lemma 7 to deduce that G is a subgraph of G_i for some $i = 1, 2, 3$.

Let $m_2(G)$ be the number of matchings in G of size two and $t(G)$ be the number of five vertex subgraphs of G comprising a degree two vertex $v \in V_2$, the two edges incident to it, and an additional isolated edge. Observe that $m_2(G_1) + t(G_1) = 6 + 0 = 6$, $m_2(G_2) + t(G_2) = 4 + 2 = 6$, and $m_2(G_3) + t(G_3) = 3 + 2 = 5$. Since $G \subset G_i$ for some $i = 1, 2, 3$, we have

$$m_2(G) + t(G) \leq \max_{i \in [3]} (m_2(G_i) + t(G_i)) = 6. \quad (3)$$

Let

$$P = \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2) \in V_1^2 \\ \mathcal{B}_1 \neq \mathcal{B}_2}} \sum_{\substack{(Y_1, Y_2) \in V_2^2 \\ Y_1 \neq Y_2}} \frac{1}{2} \chi_{\{Y_i \in \mathcal{B}_i : i=1,2\}}$$

and

$$Q = \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3) \in V_1^3 \\ \mathcal{B}_i \neq \mathcal{B}_j, i \neq j}} \sum_{\substack{(Y_1, Y_2) \in V_2^2 \\ Y_1 \neq Y_2}} \frac{1}{2} \chi_{\{Y_1 \in \mathcal{B}_1, Y_2 \in \mathcal{B}_2, Y_2 \in \mathcal{B}_3\}}.$$

Then (3) implies that

$$\mathbb{E}(P + Q) \leq 6. \quad (4)$$

By linearity of expectation, to calculate $\mathbb{E}(P+Q)$, we just need to calculate the expectations of $\chi_{\{Y_i \in \mathcal{B}_i : i=1,2\}}$ and $\chi_{\{Y_1 \in \mathcal{B}_1, Y_2 \in \mathcal{B}_2, Y_2 \in \mathcal{B}_3\}}$.

Call a pair of sets $(B_i \in \mathcal{B}_i : i = 1, 2)$ *good* if $B_1 \setminus B_2 \neq \emptyset$, $B_2 \setminus B_1 \neq \emptyset$ and $B_1 \cup B_2 \neq [n]$. Conversely, a pair is called *bad*, if either $B_1 \subset B_2$, or $B_2 \subset B_1$, or $B_1 \cup B_2 = [n]$. We see that bad pairs induce partitions on $[n]$ into at most three parts, which shows that the number of bad pairs is $O(3^n)$. Now, for each good pair (B_1, B_2) , there exists a unique partition

$$[n] = (B_1 \cap B_2) \cup (B_1 \setminus B_2) \cup (B_2 \setminus B_1) \cup ([n] \setminus (B_1 \cup B_2))$$

such that $Y_1 = B_1, Y_2 = B_2$. Therefore,

$$\begin{aligned} \mathbb{E}(\chi_{\{Y_i \in \mathcal{B}_i : i=1,2\}}) &= \mathbb{P}(Y_i \in \mathcal{B}_i : i = 1, 2) \\ &= \sum_{B_1 \in \mathcal{B}_1} \sum_{B_2 \in \mathcal{B}_2} \mathbb{P}(Y_1 = B_1, Y_2 = B_2) \\ &= \sum_{(B_i \in \mathcal{B}_i : i=1,2) \text{ is good}} \frac{1}{p(n)} \\ &= \frac{\#\{\text{good pairs } (B_i \in \mathcal{B}_i : i = 1, 2)\}}{4^n + O(3^n)} \\ &= \frac{|\mathcal{B}_1||\mathcal{B}_2| + O(3^n)}{4^n + O(3^n)} \\ &= (1 + o(1)) \frac{|\mathcal{B}_1||\mathcal{B}_2|}{4^n}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}(\chi_{\{Y_1 \in \mathcal{B}_1, Y_2 \in \mathcal{B}_2, Y_2 \in \mathcal{B}_3\}}) &= \mathbb{P}(Y_1 \in \mathcal{B}_1, Y_2 \in \mathcal{B}_2, Y_2 \in \mathcal{B}_3) \\ &= \sum_{B_1 \in \mathcal{B}_1} \sum_{B_2 \in \mathcal{B}_2 \cap \mathcal{B}_3} \mathbb{P}(Y_1 = B_1, Y_2 = B_2) \\ &= (1 + o(1)) \frac{|\mathcal{B}_1||\mathcal{B}_2 \cap \mathcal{B}_3|}{4^n}. \end{aligned}$$

Consequently, $\mathbb{E}(P + Q)$ is equal to

$$\begin{aligned}
& \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2) \in V_1^2 \\ \mathcal{B}_1 \neq \mathcal{B}_2}} \sum_{\substack{(Y_1, Y_2) \in V_2^2 \\ Y_1 \neq Y_2}} \frac{(1 + o(1))}{2} \frac{|\mathcal{B}_1||\mathcal{B}_2|}{4^n} + \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3) \in V_1^3 \\ \mathcal{B}_i \neq \mathcal{B}_j, i \neq j}} \sum_{\substack{(Y_1, Y_2) \in V_2^2 \\ Y_1 \neq Y_2}} \frac{(1 + o(1))}{2} \frac{|\mathcal{B}_1||\mathcal{B}_2 \cap \mathcal{B}_3|}{4^n} \\
&= \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2) \in V_1^2 \\ \mathcal{B}_1 \neq \mathcal{B}_2}} \frac{(6 + o(1))}{2} \frac{|\mathcal{B}_1||\mathcal{B}_2|}{4^n} + \sum_{\substack{(\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3) \in V_1^3 \\ \mathcal{B}_i \neq \mathcal{B}_j, i \neq j}} \frac{(6 + o(1))}{2} \frac{|\mathcal{B}_1||\mathcal{B}_2 \cap \mathcal{B}_3|}{4^n} \\
&= \frac{6 + o(1)}{4^n} (|\mathcal{A}_1||\mathcal{A}_2| + |\mathcal{A}_2||\mathcal{A}_3| + |\mathcal{A}_3||\mathcal{A}_1| + |\mathcal{A}_1||\mathcal{A}_2 \cap \mathcal{A}_3| + |\mathcal{A}_2||\mathcal{A}_3 \cap \mathcal{A}_1| + |\mathcal{A}_3||\mathcal{A}_1 \cap \mathcal{A}_2|).
\end{aligned}$$

Let

$$a = |\mathcal{A}_1|, b = |\mathcal{A}_2|, c = |\mathcal{A}_3|, d = |\mathcal{A}_2 \cap \mathcal{A}_3|, e = |\mathcal{A}_3 \cap \mathcal{A}_1|, f = |\mathcal{A}_1 \cap \mathcal{A}_2|.$$

It follows from (4) that

$$ab + bc + ca + ad + be + cf \leq (1 + o(1))4^n.$$

We also have by inclusion/exclusion

$$2^n \geq |\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3| = \sum_i |\mathcal{A}_i| - \sum_{i,j} |\mathcal{A}_i \cap \mathcal{A}_j|$$

which gives $a + b + c \leq d + e + f + 2^n$. Thus, we are left to solve the following optimization problem.

$$\begin{aligned}
& \max \quad abc \\
& \text{s.t.} \quad ab + bc + ca + ad + be + cf \leq (1 + o(1))4^n \\
& \quad \quad a + b + c - d - e - f \leq 2^n \\
& \quad \quad d + e \leq c, e + f \leq a, f + d \leq b \\
& \quad \quad a, b, c, d, e, f \geq 0.
\end{aligned}$$

Now, if we rescale the variables in this optimization problem by a factor of 2^n , that is, write $x' = x/2^n$ for $x \in \{a, b, c, d, e, f\}$, we get

$$\begin{aligned}
& \max \quad a'b'c' \\
& \text{s.t.} \quad a'b' + b'c' + c'a' + a'd' + b'e' + c'f' \leq 1 + \epsilon \\
& \quad \quad a' + b' + c' - d' - e' - f' \leq 1 \\
& \quad \quad d' + e' \leq c', e' + f' \leq a', f' + d' \leq b' \\
& \quad \quad a', b', c', d', e', f' \geq 0.
\end{aligned}$$

Since n can be made arbitrarily large, ϵ is arbitrarily close to zero. So we are left to show the optimization problem with $\epsilon = 0$. By solving the KKT conditions (see Appendix for the details), we get

$$\max abc \leq (0.13075 + o(1))2^{3n}.$$

This concludes the proof. \square

3 Concluding remarks

• Our basic approach is simply to average over a suitable family of partitions. It can be applied to a variety of other extremal problems, for example, it yields some results about cross intersecting families proved by Borg [5]. It also applies to the situation when the number of colors is more than the size of the forbidden configuration. In particular, the proof of Lemma 5 yields the following more general statement.

Lemma 8. *Given integers $s \geq 1$, $1 \leq t \leq k$ and $0 \leq c \leq s - 1$, let n be an integer such that $n \geq c + t(s - c)$. For $i = 1, \dots, k$, let $\mathcal{A}_i \subset \binom{[n]}{s}$ such that $\{\mathcal{A}_i\}_{i=1}^k$ contains no multicolor sunflower with t petals and core size c . Then,*

$$\sum_{i=1}^k |\mathcal{A}_i| \leq \begin{cases} \frac{(t-1)k}{m} \binom{n}{s}, & \text{if } c + t(s - c) \leq n \leq c + k(s - c) \\ (t - 1) \binom{n}{s}, & \text{if } n \geq c + k(s - c), \end{cases}$$

where $m = \lfloor (n - c)/(s - c) \rfloor$.

Note that both upper bounds can be sharp. For the first bound, when $c = 0$, $m = t < k$ and $n = ms$, let each \mathcal{A}_i consist of all s -sets omitting the element 1. A sunflower with $t = m$ petals and core size $c = 0$ is a perfect matching of $[n]$. Since every perfect matching has a set containing 1, there is no multicolor sunflower. Clearly $\sum_i |\mathcal{A}_i| = k \binom{n-1}{s} = ((t-1)k/m) \binom{n}{s}$. For the second bound, we can just take $t - 1$ copies of $\binom{[n]}{s}$ to achieve equality.

• Another general approach that applies to the sum of the sizes of families was initiated by Keevash-Saks-Sudakov-Verstraëte [16]. Both methods can be used to solve certain problems. For example, as pointed out to us by Benny Sudakov, the approach in [16] can be used to prove the $k = 3$ case of Theorem 3. On the other hand, we can use our approach to prove the following that is a very special case of a result of [16]: if we have graphs G_1, G_2, G_3 on vertex set $[n]$ with no multicolored triangle, then $|G_1| + |G_2| + |G_3| \leq 2 \binom{n}{2}$ provided $n \equiv 1, 3 \pmod{6}$. Indeed, we just take a Steiner triple system S on $[n]$ and observe (by König's theorem) that on each triple e of S , the sum over i of the number of edges of G_i within e is at most six (the result of [16] is quite a bit stronger, as it does not require the divisibility requirement and also applies when the number of colors is much larger). The same argument works for larger cliques and even for r -uniform hypergraphs, using the recent result of Keevash [15] on the existence of designs. More precisely, given integers $2 \leq r < q$ and $n > n_0$ satisfying certain divisibility conditions, if we have r -uniform hypergraphs $H_1, \dots, H_{\binom{q}{r}}$ on vertex set $[n]$ forming no multicolored K_q^r , then $\sum |H_i| \leq ((\binom{q}{r} - 1) \binom{n}{r})$ by the same proof as for triangles above except we replace a Steiner triple system with an appropriate design which is known to exist by Keevash's result.

• The main idea in the proof of Theorem 4 was to consider the graph parameter $h(G) = m_2(G) + t(G)$. The choice of $h(G)$ was obtained by a search of several parameters for which one could prove good upper bounds like in (3), while also being able to compute the expectation in (4). Perhaps some new ideas will be needed to improve the upper bound further to the conjectured value of $1/8$. Carrying out this approach for larger k appears to be difficult.

References

- [1] N. Alon, A. Shpilka, C. Umans, *On Sunflowers and Matrix Multiplication*, Computational Complexity, 22 (2013), 219–243.
- [2] C. Bey, *On Cross-Intersecting Families of Sets*, Graphs and Combinatorics, 21 (2005), 161–168.
- [3] B. Bollobás, P. Keevash, B. Sudakov, *Multicolored extremal problems*, Journal of Combinatorial Theory, Series A 107 (2004), 295–312.
- [4] P. Borg, *Intersecting and Cross-Intersecting Families of Labeled Sets*, The Electronic Journal of Combinatorics, 15, #N9, 2008.
- [5] P. Borg, *The maximum sum and the maximum product of sizes of cross-intersecting families*, European Journal of Combinatorics 35 (2014), 117–130.
- [6] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004), 243.
- [7] M. Deza, *Une Propriété Extremale des Plans Projectifs Finis Dans une Classe de Codes Equidistants*, Discrete Mathematics, 6 (1973), 343–352.
- [8] P. Erdős, C. Ko, R. Rado, *Intersection Theorems for Systems of Finite Sets*, Quarterly Journal of Mathematics: Oxford Journals, 12 (1) (1961), 313–320.
- [9] P. Erdős, R. Rado, *Intersection Theorems for Systems of Finite Sets*, Journal of London Mathematical Society, Second Series 35 (1) (1960), 85–90.
- [10] P. Erdős, E. Szemerédi, *Combinatorial Properties of Systems of Sets*, Journal of Combinatorial Theory, Series A 24 (3) (1978), 308–313.
- [11] P. Frankl, V. Rödl, *Forbidden Intersections*, Transactions of the American Mathematical Society, 300 (1) (1987), 259–286.
- [12] P. Frankl, N. Tokushige, *On r -Cross Intersecting Families of Sets*, Combinatorics, Probability and Computing 20 (2011), 749–752.
- [13] P. Frankl, S.J. Lee, M. Siggers, N. Tokushige, *An Erdős-Ko-Rado Theorem for Cross t -Intersecting Families*, Journal of Combinatorial Theory, Series A 128 (2014), 207–249.
- [14] A.J.W. Hilton, *An Intersection Theorem for a Collection of Families of Subsets of a Finite Set*, Journal of London Mathematical Society, 2 (1977), 369–384.
- [15] P. Keevash, *The existence of designs*, <http://arxiv.org/abs/1401.3665>.
- [16] P. Keevash, M. Saks, B. Sudakov, J. Verstraëte, *Multicolor Turán problems*, Advances in Applied Mathematics 33 (2004), 238–262.
- [17] P. Keevash, B. Sudakov, *Set systems with restricted cross-intersections and the minimum rank of inclusion matrices*, Siam. J. Discrete Math. Vol. 18, No. 4 (2005), 713–727.

- [18] J.H. van Lint, *A Theorem on Equidistant Codes*, Discrete Mathematics, 6 (1973), 353–358.
- [19] M. Matsumoto, N. Tokushige, *The Exact Bound in the Erdős-Rado-Ko Theorem for Cross-Intersecting Families*, Journal of Combinatorial Theory, Series A 52 (1989), 90–97.
- [20] D. Mubayi, V. Rödl, *Specified Intersections*, Transactions of the American Mathematical Society 366 (2014), No. 1, 491–504.
- [21] L. Pyber, *A New Generalization of the Erdős-Ko-Rado Theorem*, Journal of Combinatorial Theory, Series A 43 (1986), 85–90.

A The Optimization Problem in the Proof of Theorem 4

We are actually considering the following problem

$$\begin{aligned}
& \max \quad abc \\
& \text{s.t.} \quad ab + bc + ca + ad + be + cf - 1 \leq 0 \\
& \quad \quad a + b + c - d - e - f - 1 \leq 0 \\
& \quad \quad a, b, c, d, e, f \geq 0.
\end{aligned}$$

We will see that the optimal solution to this problem satisfies all the constraints in the original problem, hence solves the original problem.

We consider the KKT conditions (see, for example [6]) for this problem. Let (a, b, c, d, e, f) be an optimal solution to the problem, and $\mu_i \geq 0, i = 1, \dots, 8$ be the dual variable associated with each constraint above respectively. We have the following stationary (gradient) condition.

$$\begin{aligned}
\begin{pmatrix} bc \\ ca \\ ab \\ 0 \\ 0 \\ 0 \end{pmatrix} &= \mu_1 \begin{pmatrix} b+c+d \\ a+c+e \\ b+a+f \\ a \\ b \\ c \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&+ \mu_5 \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mu_6 \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \mu_7 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} + \mu_8 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}.
\end{aligned}$$

The complimentary slackness conditions are the following.

$$\begin{aligned}\mu_1(ab + bc + ca + ad + be + cf - 1) &= 0 \\ \mu_2(a + b + c - d - e - f - 1) &= 0 \\ \mu_3a = \mu_4b = \mu_5c = \mu_6d = \mu_7e = \mu_8f &= 0.\end{aligned}$$

First notice that clearly the maximum of abc should be positive, so in any optimal solution we have $a, b, c > 0$. By complimentary slackness conditions $\mu_3a = \mu_4b = \mu_5c = 0$, we get $\mu_3 = \mu_4 = \mu_5 = 0$.

Now, suppose that $d = e = f = 0$, then the problem is reduced to maximizing abc , subject to $ab + bc + ca \leq 1, a + b + c \leq 1$, and $a, b, c > 0$. It is easy to see that in this case $a = b = c = 1/3$ solves the problem, with the maximum $1/27$. Therefore, we may assume without loss of generality that $d > 0$, so by complimentary slackness, $\mu_6 = 0$. Thus, we have reduced the stationary condition into the following form.

$$\begin{pmatrix} bc \\ ca \\ ab \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mu_1 \begin{pmatrix} b + c + d \\ a + c + e \\ b + a + f \\ a \\ b \\ c \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\mu_7 \\ -\mu_8 \end{pmatrix}.$$

If $\mu_1 = 0$, then we get that $\mu_2 = 0$ also, then $a = b = c = 0$, a contradiction. This shows that $\mu_1 > 0$. Now suppose that $\mu_2 = 0$. Then we get $a = \mu_2/\mu_1 = 0$ which makes the maximum also 0, so we must have $\mu_2 > 0$. By complimentary slackness, the first two constraints both hold with equality. Next, we use the stationary condition to express a, b, c in terms of μ_1, μ_2, μ_7 and μ_8 .

$$a = \frac{\mu_2}{\mu_1}, \quad b = \frac{\mu_2 + \mu_7}{\mu_1}, \quad c = \frac{\mu_2 + \mu_8}{\mu_1}. \quad (5)$$

Case 1. $\mu_7 = \mu_8 = 0$. We get $a = b = c = \mu_2/\mu_1$. In this case, we are solving the maximization of a^3 , subject to $3a^2 + ax = 1, 3a - x = 1$, and $a, x > 0$, where $x = d + e + f$. The optimality is obtained at $a = x = 1/2$, with maximum $1/8$.

Case 2. Exactly one of μ_7 and μ_8 is positive. Without loss of generality, we may assume $\mu_7 > 0$, which implies that $e = 0$. Since in this case $\mu_8 = 0$, we get $a = c$ from (5), so the problem is reduced to

$$\begin{aligned}\max \quad & a^2b \\ \text{s.t.} \quad & 2ab + a^2 + ax - 1 = 0 \\ & 2a + b - x - 1 = 0 \\ & a, b, x > 0\end{aligned}$$

where $x = d + f$. This problem has the maximum $\frac{1}{729} (29 + 20\sqrt{10}) \approx 0.126537$ at

$$a = \frac{1}{9} (1 + \sqrt{10}) \approx 0.462475, \quad b = -\frac{-29 - 20\sqrt{10}}{9 (1 + \sqrt{10})^2} \approx 0.591617,$$

and

$$x = -\frac{81 \left(-\frac{1}{729} 2 (1 + \sqrt{10})^3 + \frac{1}{81} (1 + \sqrt{10})^2 + \frac{1}{729} (-29 - 20\sqrt{10}) \right)}{(1 + \sqrt{10})^2} \approx 0.516568.$$

Case 3. Both $\mu_7, \mu_8 > 0$. This implies that $e = f = 0$. So from the stationary condition, we get

$$ac = \mu_1(a + c) + \mu_2$$

$$ab = \mu_1(a + b) + \mu_2.$$

By (5), we can eliminate the variables a, b, c to get

$$\mu_2(\mu_2 + \mu_8) = \mu_1^2(\mu_2 + (\mu_2 + \mu_8)) + \mu_1^2\mu_2$$

$$\mu_2(\mu_2 + \mu_7) = \mu_1^2(\mu_2 + (\mu_2 + \mu_7)) + \mu_1^2\mu_2.$$

A bit of algebra shows that $\mu_7 = \mu_8$, which implies $b = c$. Thus the problem is reduced to

$$\begin{aligned} \max \quad & ab^2 \\ \text{s.t.} \quad & 2ab + b^2 + ad - 1 = 0 \\ & a + 2b - d - 1 = 0 \\ & a, b, d > 0 \end{aligned}$$

Similar to the previous cases, one can solve system using the method of Lagrange multipliers as follows. We first eliminate variable d to get $2ab + b^2 + a(a + 2b - 1) - 1 = 0$. Then define the Lagrangian as

$$L(a, b, \lambda) = ab^2 + \lambda(a^2 + 4ab + b^2 - a - 1).$$

To find the maximum, we need to solve the following system of equations.

$$\begin{aligned} \frac{\partial L}{\partial a} &= b^2 + 2\lambda a + 4\lambda b - \lambda = 0 \\ \frac{\partial L}{\partial b} &= 2ab + 4\lambda a + 2\lambda b = 0 \\ \frac{\partial L}{\partial \lambda} &= a^2 + 4ab + b^2 - a - 1 = 0. \end{aligned}$$

One can obtain exact closed form solutions to this system involving radicals, however, the formulas are too long to display, so we give the approximation as follows. The maximum is approximately equal to 0.130748, with $a \approx 0.37478, b \approx 0.590649, \lambda \approx -0.165171$. This corresponds to the following solution to the original problem:

$$a \approx 0.37478, \quad b = c \approx 0.590649, \quad d \approx 0.556078, \quad e = f = 0.$$

Comparing with all the other cases, this is the actual maximum that abc can achieve. Moreover, this solution satisfies the constraints $d + e \leq c$, $e + f \leq a$ and $f + d \leq b$, hence is the optimal solution to the original problem in the proof of Theorem 4 as well.